PROTEIN STRUCTURE REPORT

The thermo- and acido-stable ORF-99 from the archaeal virus AFV1

Adeline Goulet,[1] Silvia Spinelli,[1] Stéphanie Blangy,[1] Herman van Tilbeurgh,[2] Nicolas Leulliot,[2] Tamara Basta,[3] David Prangishvili,[3] Christian Cambillau,[1]* and Valérie Campanacci[1]*

[1]Architecture et Fonction des Macromolécules Biologiques, CNRS and Universités d'Aix-Marseille I & II, France
[2]Institut de Biochimie et de Biophysique Moléculaire et Cellulaire (CNRS-UMR 8619), France
[3]Institut Pasteur, Unité de Biologie Moléculaire du Gène chez les Extrêmophiles, France

Received 24 February 2009; Revised 9 March 2009; Accepted 16 March 2009
DOI: 10.1002/pro.122
Published online 30 March 2009 proteinscience.org

Abstract: Acidianus Filamentous Virus 1 (AFV1), isolated from acidic hot springs, is an enveloped lipid-containing archaeal filamentous virus with a linear double-stranded DNA genome. It infects *Acidianus*, which is a hyperthermostable archaea growing at 85°C and acidic pHs, below pH 3. AFV1-99, a protein of 99 amino acids of unknown function, has homologues in the archaeal virus families Lipothrixviridae and Rudiviridae. We determined the crystal structure of AFV1-99 at 2.05 Å resolution. AFV1-99 has a new fold, is hyperthermostable (up to 95°C) and resists to extreme pH (between pH 0 and 11) and to the combination of high temperature (95°C) and low pH (pH 0). It possesses characteristics of hyperthermostable proteins, such as a high content of charged residues.

Keywords: crenarchaea; hyperthermostability; acidophile

Introduction

Double-stranded DNA (dsDNA) viruses that infect Archaea living in acidic hot springs (temperatures above 80°C and pH below 3) are radically different in their properties from viruses that infect Bacteria and Eukarya. Not only are the shapes of these viruses distinct from those of all other viruses found on Earth to-day, but also ~80% of their ORFs do not share any sequence homology with ORFs of other viruses or of cellular life forms, apart from other archaeal viruses.[1,2] Filamentous viruses, the most abundant morphotype in these extreme environments, form the new viral order Ligamenvirales divided into Rudiviridae and Lipothrixviridae families. Lipothrixviruses (AFV1-8, TTV1-3, and SIFV)[1] infect acidophilic and hyperthermophilic Crenarchaea from the

Additional Supporting Information may be found in the online version of this article.

Abbreviations: AFV1, *Acidianus* filamentous virus 1; ORF, open reading frame; CD, circular dichroism; R.m.s.d., root mean square deviation; TLS, Translation/Libration/Screw; NCS, non-crystallographic symmetry.

Grant sponsor: Marseille-Nice Génopole®, French Research Ministry; Grant sponsor: Agence Nationale de la Recherche (VIRAR); Grant number: NT05-2_41674.

**Correspondence to:* Valérie Campanacci, AFMB, UMR 6098, Case 932, 163 avenue de Luminy, 13288 Marseille cedex 9, France. E-mail: valerie.campanacci@afmb.univ-mrs.fr or Christian Cambillau, AFMB, UMR 6098, Case 932, 163 avenue de Luminy, 13288 Marseille cedex 9, France.
E-mail: cambillau@afmb.univ-mrs.fr

genera Acidianus or Sulfolobus and, are the first enveloped lipid-containing filamentous viruses with linear dsDNA genomes to be discovered. The lipothrixvirus Acidianus Filamentous Virus 1 (AFV1) has been isolated from a hot spring in Yellowstone National Park and infects acidophilic and hyperthermophilic members of *Acidianus* genus.[3] Its 20.8 kb linear ds-DNA genome encodes for ~40 ORFs with 32% of them homologous to viral ORFs from the lipothrixvirus SIFV and the Rudiviruses SIRV1 and SIRV2, and with few predicted functions based on sequence comparisons, two glycosyltransferases, two CopG like proteins and one transcription regulator. Apart from structural proteins, it is expected that other AFV1 ORFs should code for enzymes or proteins involved in viral infection and replication, and have functional equivalents in bacteria, archaea or eukarya. Structural and functional data on archaeal virus proteins are still extremely scarce but are needed to better understand their evolution and biology. Eight protein structures from archaeal viruses have been reported, three from the spindle-shaped SSV1 (Fuselloviridae),[4–6] four from the unclassified icosahedral STIV,[7–10] and one from AFV3 (Lipothrixviridae).[11] To get insight into the function of the Lipothrixviridae proteome, we have initiated a program to determine the structure of several conserved ORFs.[12] Moreover, they may reveal novel features that can be exploited for biotechnological applications.

We present here the crystal structure of AFV1-99 (Lipothrixviridae), a protein of 99 residues that has homologues in other filamentous archaeal viruses, *Sulfolobus islandicus* rod-shaped virus 1 (SIRV1, Rudiviridae; ORF99, ORF100, ORF95 and ORF96) and *Sulfolobus islandicus* filamentous virus (SIFV1, Lipothrixviridae; ORF14). The fold of AFV1-99 is novel and it is extremely thermostable and also resistant to very low pH (pH 0).

Results and Discussion

Structures of the native and SeMet AFV1-99

Crystals of native AFV1-99 belong to space group $P4_12_12$ ($a = b = 101.4$ Å, $c = 58.5$ Å) and diffract weakly at 3.1 Å resolution. Because AFV1-99 contains neither methionines nor cysteines, Leu27 and Ile70 predicted to be in an α-helix and in a β-strand, respectively, were mutated into methionine for phasing. The SeMet substituted double mutant protein was produced in *E. coli*, purified, and the incorporation of SeMet was checked by mass spectrometry. The SeMet protein crystallized in a space group different to that of the native protein ($P2_12_12_1$, $a = 57.1$ Å, $b = 97.7$ Å, $c = 103.7$ Å) and diffracted to 2.05 Å resolution. The structure was solved by SAD at the SeMet anomalous peak. The data were of sufficient quality to locate the 8 Se sites present in the four molecules in the asymmetric unit. The current model is refined to 2.05 Å, resulting in an R_{factor} of 0.19 and an R_{free} of 0.22 (Table I). The tetramer can be described as a dimer of dimers, each one assembled through a tight interface. It should be noticed that the protein is purified as a dimer (not shown).

The structure of the native protein was solved by molecular replacement using one monomer of the SeMet protein as search model. Two monomers were localized in the asymmetric unit and were refined to 3.1 Å with a final R_{factor} of 0.245 and an R_{free} of 0.296 (Table I). This dimer is identical to the tight dimer observed in the SeMet labeled structure (r.m.s.d. = 0.45 Å). It buries 836 Å^2 (12.8%) of the accessible surface, involves 21% of the residues and is made up of an anti-parallel arrangement of two β-strands and two α-helices from two monomers. Crystal packing reconstitutes the same tetrameric organization observed in

Table I. *Data collection, Phasing, and Refinement Statistics*

AFV1-99	SeMet	Native
Data collection		
Space group	$P2_12_12_1$	$P4_12_12$
Cell dimensions a, b, c (Å)	$a = 57.1, b = 97.7, c = 103.7$	$a = b = 101.4\ c = 58.5$
Wavelength (Å)	Peak (SAD) 0.979	0.934
Resolution (Å)[a]	45–2.05 (2.16–2.05)	40–3.1 (3.27–3.1)
R_{sym}[a,b]	0.095 (0.39)	0.10 (0.56)
I/σI[a]	25 (7)	15 (3.3)
Completeness (%)[a]	100 (100)	99.7 (100)
Redundancy[a]	13.8 (14.1)	7.5 (7.8)
Refinement		
Resolution (Å)[a]	30–2.05 (2.10–2.05)	15–3.1 (3.17–3.10)
No. reflections	35,982	5246 (361)
Completeness (%)[a]	100 (100)	99.7 (100)
R_{work} / R_{free}[a,c]	0.19/0.22 (0.191/0.23)	0.24/0.29 (0.31/0.37)
No. atoms	3740	1582
B-factor (overall, Å^2)	18.34	66.8
rmsd bond lengths (Å), bond angles (°)	0.013, 1.44	0.011, 0.009

[a] Values in parentheses refer to the highest-resolution bin.
[b] Rsym = $\Sigma_h\Sigma_i|I_{hi} - (I_h)|/\ \Sigma_h\Sigma I(I_h)$, where I_i is the ith observation of reflection h and (I_h) is the weighted average intensity for all observations l of reflection h.
[c] Calculated from randomly choosen reflections (5%).

the SeMet structure: both "tetramers" structures superimpose with a r.m.s.d. of 1.16 Å. The inter-dimeric surface is smaller (379 Å^2, 6% of the total surface). It is stabilized by antiparallel interactions between β-strands 1a, residues 46 to 50, by hydrogen bonds between side-chain residues, and by nonsymmetrical hydrogen bonds. In contrast with the first tight interface, no hydrophobic interactions are established. The SeMet labeled structure being solved at higher resolution, the rest of the structural description will be based on it.

AFV1-99 structure: a new α/β fold

The AFV1-99 structure is made up of two helices and a β-sheet of five anti-parallel β-strands (see Fig. 1). Proceeding from the N- to the C-terminus, we find successively α-helix 1 (1–19), β-strand 1 (23–33) split in two parts (1a and 1b, Fig. 1), a short helix 2 (35–42) followed by an extended stretch; residues 46 to 96 form the others 4 antiparallel β-strands. Both helices cover one side of the β-sheet whereas the other side is exposed to solvent. Searching with DALI[13] for similar proteins in the PDB produced no significant hits, with low *Z* values (below 5.4) and relatively high r.m.s.d. values (2.5 Å at best) on a small number of residues (~60). All these hits were essentially overlapping with the β-sheet (strands 2-5). AFV1-99 structure defines thus a new α/β fold that is conserved in crenarchaeal linear Lipothrixviruses and Rudiviruses and identified as the Pfam071118 domain, DUF1374 superfamily.

Because of the novelty of the fold, the function of AFV1-99 cannot be assigned directly. AFV1-99 does not exhibit any crevice or cleft able to bind a small molecule and hence to perform catalysis. However, AFV1-99 tetramers (see above) are assembled in long fibers with good interacting surfaces in the crystal packing, which may point to a role as structural protein.

AFV1-99 is a hyperthermostable and hyperacidostable protein

AFV1 viruses proliferate in media at extreme pH values (<1.5). Structural proteins have therefore to sustain fold and function not only at high temperature but also at low pH. The AFV1-99 CD spectrum recorded at 20°C and pH 7.7 displays a single broad minimum at 221 nm and a strong positive band at 197 nm characteristic of a folded α/β protein containing mainly β-strands. The CD spectra recorded at 20°C and 95°C [Supporting Information Fig. 2(a)] and at neutral (7.7) and acidic (0) pH conditions [Supporting Information Fig. 2(b)] are all perfectly superimposable establishing its extreme temperature and pH resistance. Furthermore, it is also stable, but to a lesser extent, to high pH values up to pH 11 [Supporting Information Fig. 2(b)]. Not only AFV1-99 resists to very low pH, but also resists to combination of low pH and high temperature. The CD spectrum recorded at pH 0 and 95°C suggests it remains folded [Supporting Information Fig. 2(c)]. In contrast, at pH 11 and 95°C, AFV1-99 becomes unfolded: the signal at 200 nm is negative and the signal in the range 215–260 nm is closer to zero [Supporting Information Fig. 2(d)].

Over the past decades proteins from hyperthermophiles have been extensively studied in order to identify general rules governing the molecular basis of protein thermostability.[14–18] Although these issues are still debated, two prominent observations seem to be generally associated with thermostability: structural compactness and conformational rigidity. AFV1-99 is endowed with characteristics favorable to thermal stability: a very compact α/β fold, structured N and C

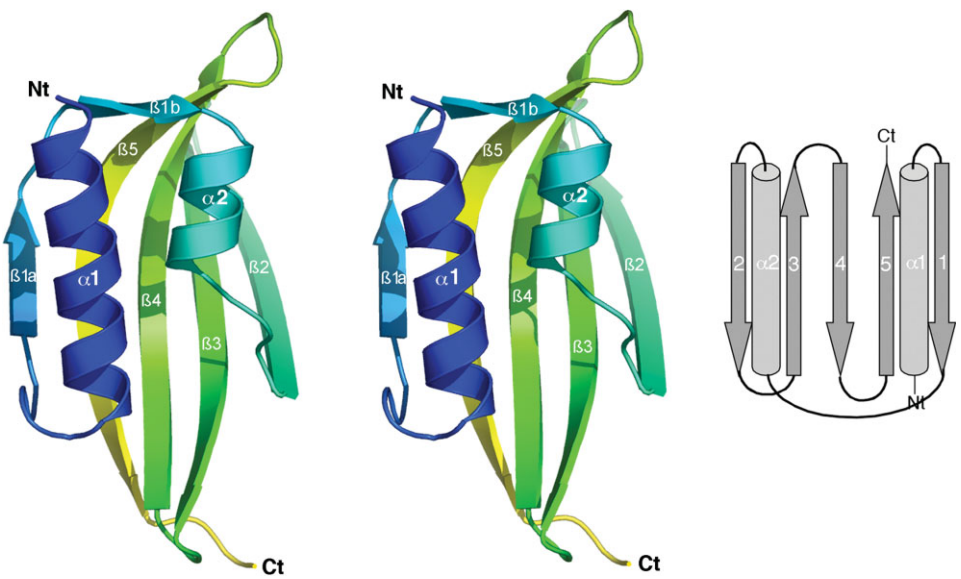

Figure 1. A: Stereo view of the 3D structure of AFV1-99. Colors vary from blue (N-terminus) to yellow (C-terminus). Helices and β-strands are numbered from 1 to 2 and 1 to 5, respectively. B: Topological diagram of AFV1-99 fold.

termini, short loops[19] and the absence of cavities. Thermostability also seems to be correlated with biased amino acid content and spatial distribution of residue types that may establish favorable noncovalent interactions at high temperatures.[18] A comparative proteome analysis using 30 complete genome sequences from the three kingdoms clearly demonstrated that a higher proportion of charged (Lys and Glu) versus polar (noncharged) amino acids, is a signature of hyperthermophily.[18] With 34.4% charged residues (Ch) versus 17.2% of polar residues (Po), the sequence of AFV1-99 corresponds to these criteria. This difference (Ch-Po) of 17.2% is larger than the 10% difference threshold observed in the extreme thermophilic orfeomes. Besides, AFV1-99 contains many acidic residues with a high ratio of acidic to basic residues. Acidic residues map around the surface and half of them interact directly with main and side chain groups. Charged and polar residues at the surface increase the density of hydrogen bonding with the surrounding water, which is a feature correlated with thermostabilty.[20] We also postulate that in the case of AFV1-99 this proportion and repartition of acidic residues could explain its stability at low pH. Indeed, even at pH inferior to their pKas, protonated carbonyl groups should still be able to form the hydrogen bonds network observed in the structure.

Materials and Methods

Native AFV1-99 expression, purification, and characterization

Native AFV1-99 was cloned into the Gateway pDEST17 expression vector (Invitrogen). A sequence encoding a TEV protease cleavage site was inserted between the attB1 and the gene of interest. AFV1-99 was overexpressed in *E. coli* Rosetta(DE3)pLysS strain (Novagen) grown in Terrific Broth (Gibco) at 25°C. Overexpressed proteins were purified in three steps on a Pharmacia Akta FPLC using (i) nickel affinity chromatography (His-Trap 5 mL column, GE Healthcare) (ii) a second nickel affinity chromatography, after cleavage of the N-terminal hexahistidine tag with TEV protease and (iii) and a preparative Superdex 200 HR26/60 gel filtration in 10 m*M* CHES pH 9, NaCl 150 m*M*. Protein was concentrated using Amicon Ultra-15 mL 5K (Millipore) and characterized by SDS-PAGE, MALDI-TOFF mass spectrometry (Brüker Autoflex). The molecular weight was determined using an analytical size exclusion chromatography (SEC) on a HPLC system (Waters) with on-line multi-angle laser light-scattering absorbance and refractive index detectors (MALLS/UV/RI) (Wyatt technology). SEC was performed on a Superose12 10/30 column (Amersham Biosciences) in the eluent Tris 10 m*M*, pH 8, NaCl 300 m*M* and the molecular weight was calculated by the ASTRA V software (Wyatt technology) using a d*n*/d*c* of 0.180 mL/g.

Double methionine mutant production for selenomethionine-labeling

Double-amino-acid mutations of AFV1-99, Leu27Met and Ile70Met, were generated using the QuickChange Multi site-directed mutagenesis kit (Stratagene). Both substitutions were performed consecutively, first Leu27Met substitution and then Ile70Met. Parental plasmids were digested by *DpnI*, and the vectors containing the mutation were transformed in XL10-Gold ultracompetent cells. Transformants were selected onto LB ampicillin (100 μg/mL) agar plates. Mutations were verified by automated DNA sequencing (Genome Express, France).

SelenoMethionine (SeMet) labeled mutant was prepared following standard procedures in minimum medium M9 by blocking the methionine biosynthesis pathway.[21] Expression, purification, and characterization of the SeMet labeled AFV1-99 were carried out using the same protocols as for native protein.

Stability studies to temperature and extreme pHs

Temperature and pH stability studies were carried out by far-UV Circular Dichroism (CD) spectroscopy. CD spectra were recorded with a JASCO J-810 spectropolarimeter (JASCO Corporation, Japan) equipped with a Peltier temperature control system. Far-UV measurements (185–260 nm) were performed using a 0.1 cm path quartz cuvette, with a scanning speed of 20 nm/min, spectral bandwidth of 1 nm, and were averaged over three scans. The solvent spectra were subtracted in all experiments to eliminate background effects. CD measurements in millidegrees were performed at a protein concentration of 0.1 mg/mL in 10 m*M* Na/Na_2 phosphate buffer pH 7.7. Stability tests at pH 0 and pH 11 were carried out in HCl 1*M* and NaOH 0.001*M*, respectively. Thermal denaturation was monitored by increasing temperature from 20°C to 95°C with a heating rate of 5°C/min recording CD signal at 221 nm and CD spectra at 95°C.

Crystallization and structure determination

SeMet labeled AFV1-99 was crystallized in Greiner plates in condition 7 (40% 2-methyl-2,4-pentanediol, 0.2*M* calcium chloride) from the commercially available screen MPD (Hampton Research) using a Cartesian nanodrop-dispensing robot. After optimization, diffraction-quality crystal were obtained in sitting drops by mixing 200 nL of protein at 8.4 mg/mL with 100 nL of 33% 2-methyl-2,4-pentanediol (MPD) (v/v), 0.2*M* calcium chloride 0.1*M* HEPES pH 6.8. Crystals were flash-frozen in liquid nitrogen for data collection at 100 K. One Single-wavelength Anomalous Diffraction (SAD) dataset was collected at the Se-K edge (0.979 Å) at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) beamline ID29 using a ADSC Q315r detector. Diffraction data were processed using MOSFLM/SCALA (see Table I).[22] Selenium-substructure (8 sites)

was solved using SHELXD, phases calculations and density modification were performed using SHELXE.[23] ArpWarp[24] was used to autotrace ~90% of the asymmetric unit, the remaining part was manually built using COOT.[25] The model was refined with REFMAC5[26] at 2.05 Å using TLS (Translation/Libration/Screw) parameters applied to four TLS groups, and applying restrained non-crystallographic symmetry (NCS) during the first few rounds of refinement.

Native AFV1-99 was crystallized in Greiner plates in condition 29 (1*M* sodium citrate, 0.1*M* Sodium HEPES pH 7.5) from the commercially available screen MDL (Molecular Dimension Limited). After optimization, diffraction-quality crystal was obtained in Linbro plates by mixing 1 μL of protein at 11.5 mg/mL with 1 μL of 1.35*M* sodium citrate, 0.1*M* Sodium HEPES pH 8.0. One native dataset was collected at the ESRF beamline ID14-eh1 using a MAR CCD detector. The structure of the native protein was solved by molecular replacement with MOLREP[26] using SeMet labeled AFV1-99 as initial search model (3DF6, chain A). Rigid body and restrained refinement at 3.1 Å, applying restrained NCS and 4 TLS groups per chain and TLS Motion Determination,[27] converged readily to a good model with R and R_{free} of 0.245 and 0.296, respectively (Table I). The quality of the final models was checked with PROCHECK.[28] Atomic coordinates and structure factors of native and SeMet labeled AFV1-99 have been deposited at the Protein Data Bank at RCSB as entries 3DJW and 3DF6, respectively. The figures were produced with Pymol, http://www.pymol.org.

References

1. Prangishvili D, Forterre P, Garrett RA (2006) Viruses of the Archaea: a unifying view. Nat Rev Microbiol 4:837–848.
2. Prangishvili D, Garrett RA, Koonin EV (2006) Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. Virus Res 117:52–67.
3. Bettstetter M, Peng X, Garrett RA, Prangishvili D (2003) AFV1, a novel virus infecting hyperthermophilic archaea of the genus *Acidianus*. Virology 315:68–79.
4. Kraft P, Kummel D, Oeckinghaus A, Gauss GH, Wiedenheft B, Young M, Lawrence CM (2004) Structure of D-63 from *Sulfolobus* spindle-shaped virus 1: surface properties of the dimeric four-helix bundle suggest an adaptor protein function. J Virol 78:7438–7442.
5. Menon SK, Maaty WS, Corn GJ, Kwok SC, Eilers BJ, Kraft P, Gillitzer E, Young MJ, Bothner B, Lawrence CM (2008) Cysteine usage in *Sulfolobus* spindle-shaped virus 1 and extension to hyperthermophilic viruses in general. Virology 376:270–278.
6. Kraft P, Oeckinghaus A, Kummel D, Gauss GH, Gilmore J, Wiedenheft B, Young M, Lawrence CM (2004) Crystal structure of F-93 from *Sulfolobus* spindle-shaped virus 1, a winged-helix DNA binding protein. J Virol 78:11544–11550.
7. Larson ET, Eilers BJ, Reiter D, Ortmann AC, Young MJ, Lawrence CM (2007) A new DNA binding protein highly conserved in diverse crenarchaeal viruses. Virology 363:387–396.
8. Larson ET, Eilers B, Menon S, Reiter D, Ortmann A, Young MJ, Lawrence CM (2007) A winged-helix protein from *Sulfolobus* turreted icosahedral virus points toward stabilizing disulfide bonds in the intracellular proteins of a hyperthermophilic virus. Virology 368:249–261.
9. Larson ET, Reiter D, Young M, Lawrence CM (2006) Structure of A197 from *Sulfolobus* turreted icosahedral virus: a crenarchaeal viral glycosyltransferase exhibiting the GT-A fold. J Virol 80:7636–7644.
10. Khayat R, Tang L, Larson ET, Lawrence CM, Young M, Johnson JE (2005) Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses. Proc Natl Acad Sci USA 102:18944–18949.
11. Keller J, Leulliot N, Cambillau C, Campanacci V, Porciero S, Prangishvilli D, Forterre P, Cortez D, Quevillon-Cheruel S, van Tilbeurgh H (2007) Crystal structure of AFV3-109, a highly conserved protein from crenarchaeal viruses. Virol J 4:12–22.
12. Goulet A, Spinelli S, Campanacci V, Porciero S, Blangy S, Garrett RA, van Tilbeurgh H, Leulliot N, Basta T, Prangishvili D, Cambillau C (2006) Crystallization and preliminary X-ray diffraction analysis of protein 14 from *Sulfolobus islandicus* filamentous virus (SIFV). Acta Crystallogr Sect F Struct Biol Cryst Commun 62:884–886.
13. Holm L, Sander C (1995) Dali: a network tool for protein structure comparison. Trends Biochem Sci 20:478–480.
14. Vieille C, Zeikus GJ (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol Mol Biol Rev 65:1–43.
15. Li WF, Zhou XX, Lu P (2005) Structural features of thermozymes. Biotechnol Adv 23:271–281.
16. Chakravarty S, Varadarajan R (2002) Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study. Biochemistry 41:8152–8161.
17. Unsworth LD, van der Oost J, Koutsopoulos S (2007) Hyperthermophilic enzymes--stability, activity and implementation strategies for high temperature applications. FEBS J 274:4044–4056.
18. Cambillau C, Claverie JM (2000) Structural and genomic correlates of hyperthermostability. J Biol Chem 275:32383–32386.
19. Thompson MJ, Eisenberg D (1999) Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. J Mol Biol 290:595–604.
20. Vogt G, Woell S, Argos P (1997) Protein thermal stability, hydrogen bonds, and ion pairs. J Mol Biol 269:631–643.
21. Doublie S (1997) Preparation of selenomethionyl proteins for phase determination. Methods Enzymol 276:523–530.
22. Collaborative Computational Project, number 4, 4 C. C. P. N (1994) The CCP4 suite: programs for crystallography. Acta Crystallogr D Biol Crystallogr 50:760–766.
23. Sheldrick GM (2008) A short history of SHELX. Acta Crystallogr A 64, 112–122.
24. Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. Nat Struct Biol 6:458–463.
25. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60:2126–2132.
26. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53:240–255.
27. Painter J, Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr 62:439–450.
28. Laskowski R, MacArthur M, Moss D, Thornton J (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallog 26:91–97.